The average time to first flower (measured in growing degree days, or GDD) in *Acer saccarhum* is 25.2. We want to know if a related species, *Acer rubrum* has the same flowering time.

Step 1:
H_0: The mean flowering time in *Acer rubrum*, $\mu_r = 25.2$

H_A: $\mu_r \neq 25.2$

Step 2: Random sample:

plant #	time to flower (GDD)
1	26.3
2	27.5
3	25.3
4	29.1
5	28.7

Step 3: $n = 5$ $\bar{Y} = 27.38$ $s = 1.597$

Steps 4 of hypothesis testing: Determine P-value:

1. choose a test statistic.
2. What values of the test statistic are expected under H_0?
3. How does the observed test statistic differ from these expectations?
4. What is the probability of observing a value of a test statistic this extreme or more extreme if the H_0 is true? – this is the P-value.

$$Z = \frac{\bar{Y} - \mu_0}{\left(\frac{\sigma_0}{\sqrt{n}}\right)}$$

The null distribution of z is the standard normal:

$$Z \sim \text{normal}(\mu = 0, \sigma = 1)$$

$$t = \frac{\bar{Y} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$$

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

The null distribution of t is Student's t-distribution with $n - 1$ degrees of freedom:

$$T \sim t\text{-distribution}(df = n - 1)$$

$$t = \frac{\bar{Y} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{27.38 - 25.2}{\left(\frac{1.597}{\sqrt{5}}\right)}$$

$$t = 3.05$$

Critical values:

$$t_{0.05(2),4} = 2.78$$

Step 5: $t >$ the critical value. P-value < 0.05. Reject the H_0.

Step 6: A sample of flowering time from 5 randomly selected *Acer rubrum* individuals had a mean flowering time of 27.38 GDD ($s = 1.597$). This mean is larger than the mean in *Acer saccarhum* (25.2GDD), and the difference is too large to be explained by sampling error ($P < 0.05$, and $t = 3.05$ with $df = 4$).

Critical values:

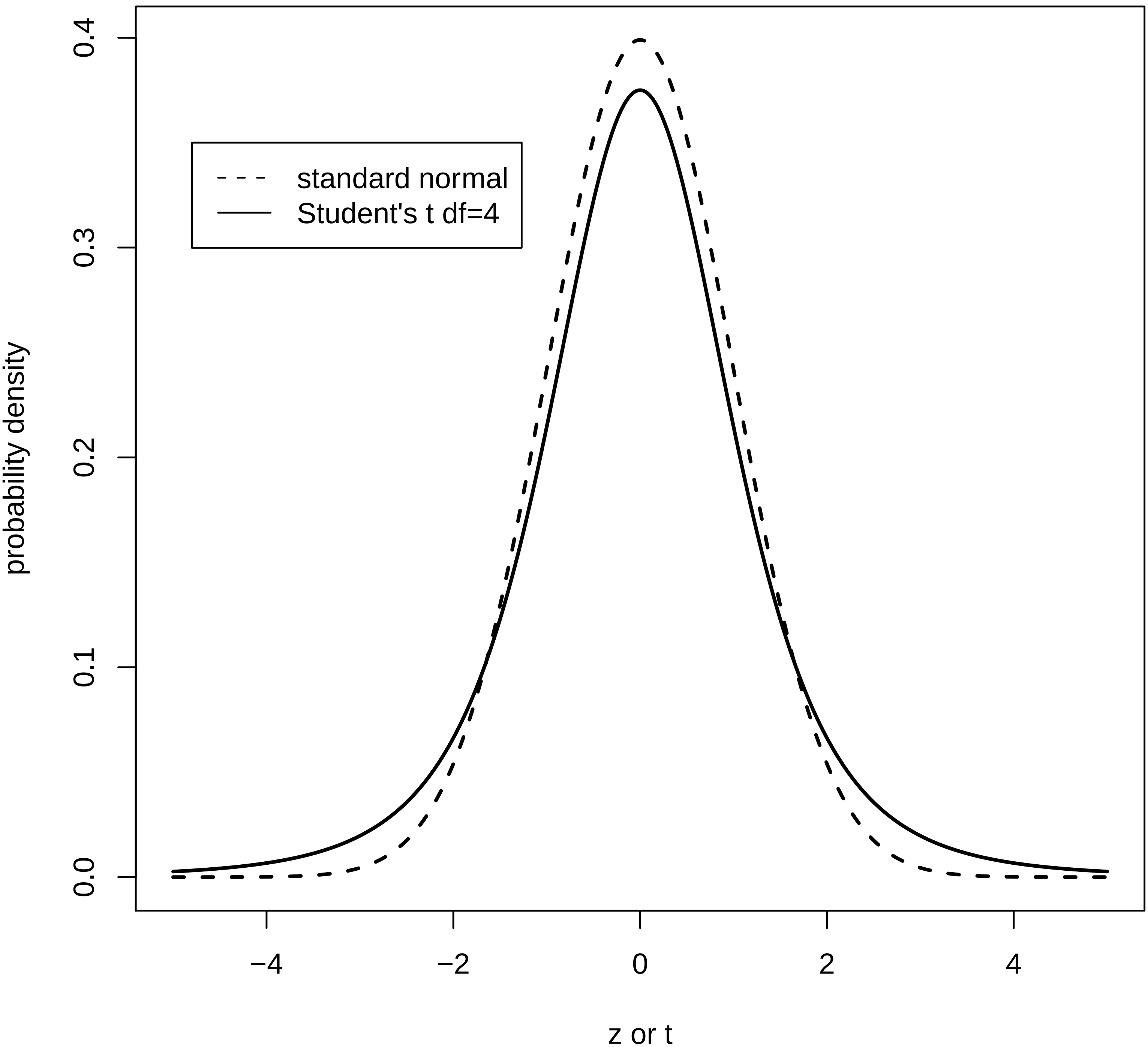
$$t_{0.05(2),4} = 2.78$$

Notice this is larger than the critical value from the standard normal:

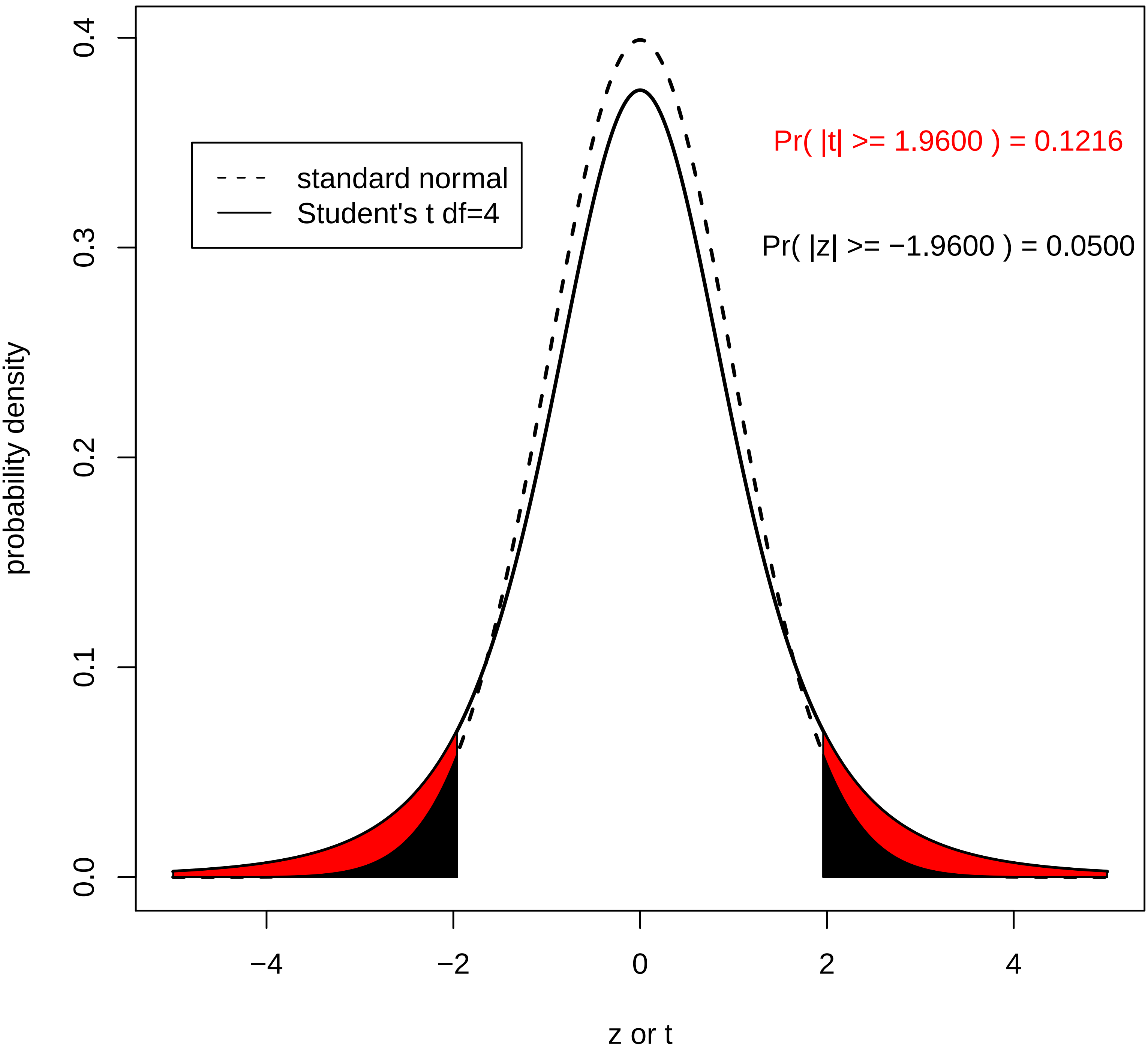
$$Z_{0.05(2)} = 1.96$$

It makes sense that it is "harder" to reject when using a t-test compared to a z test:

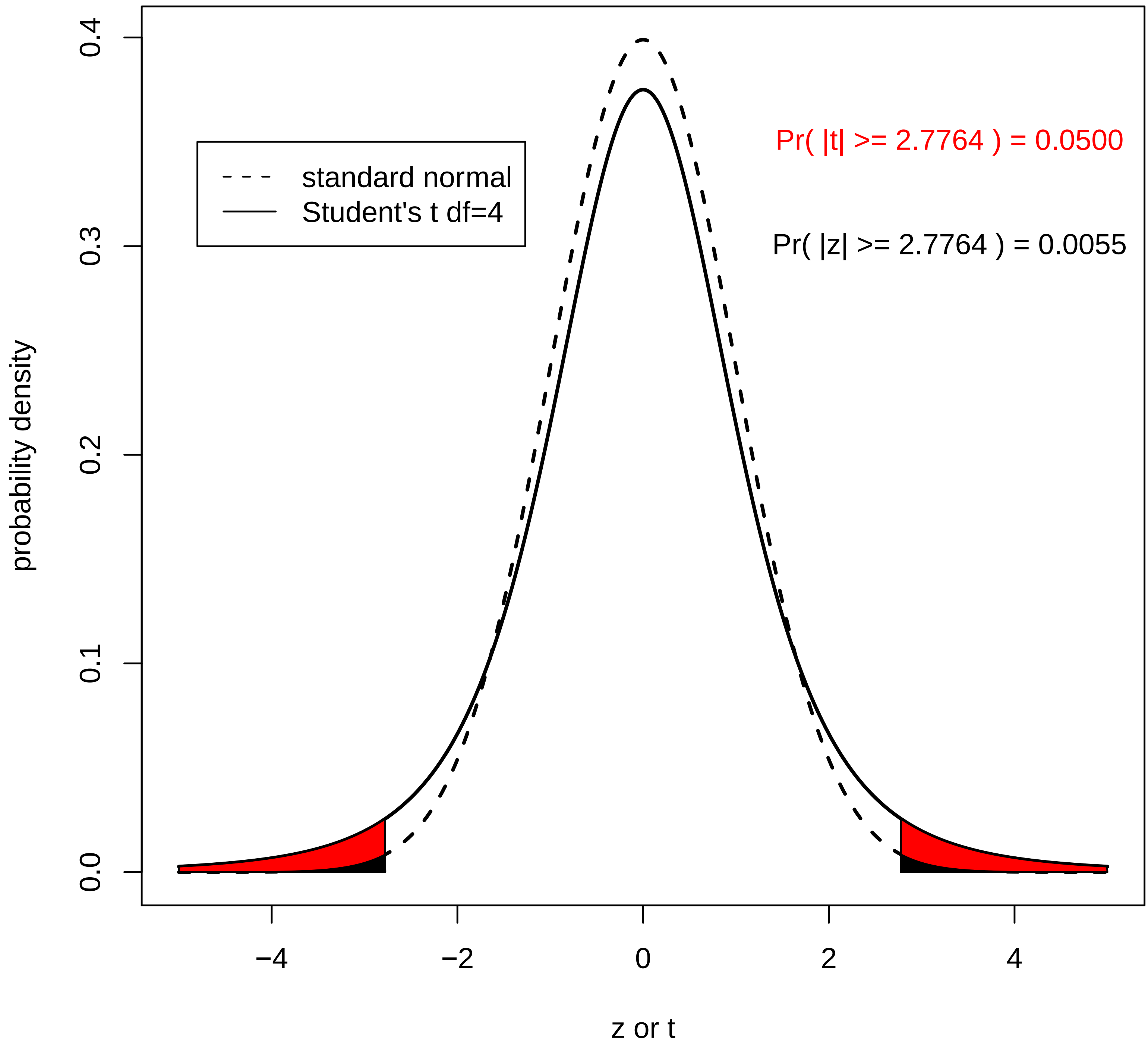
When performing a t-test, we have to consider the possibility that our test statistic is high because we have underestimated the standard deviation.

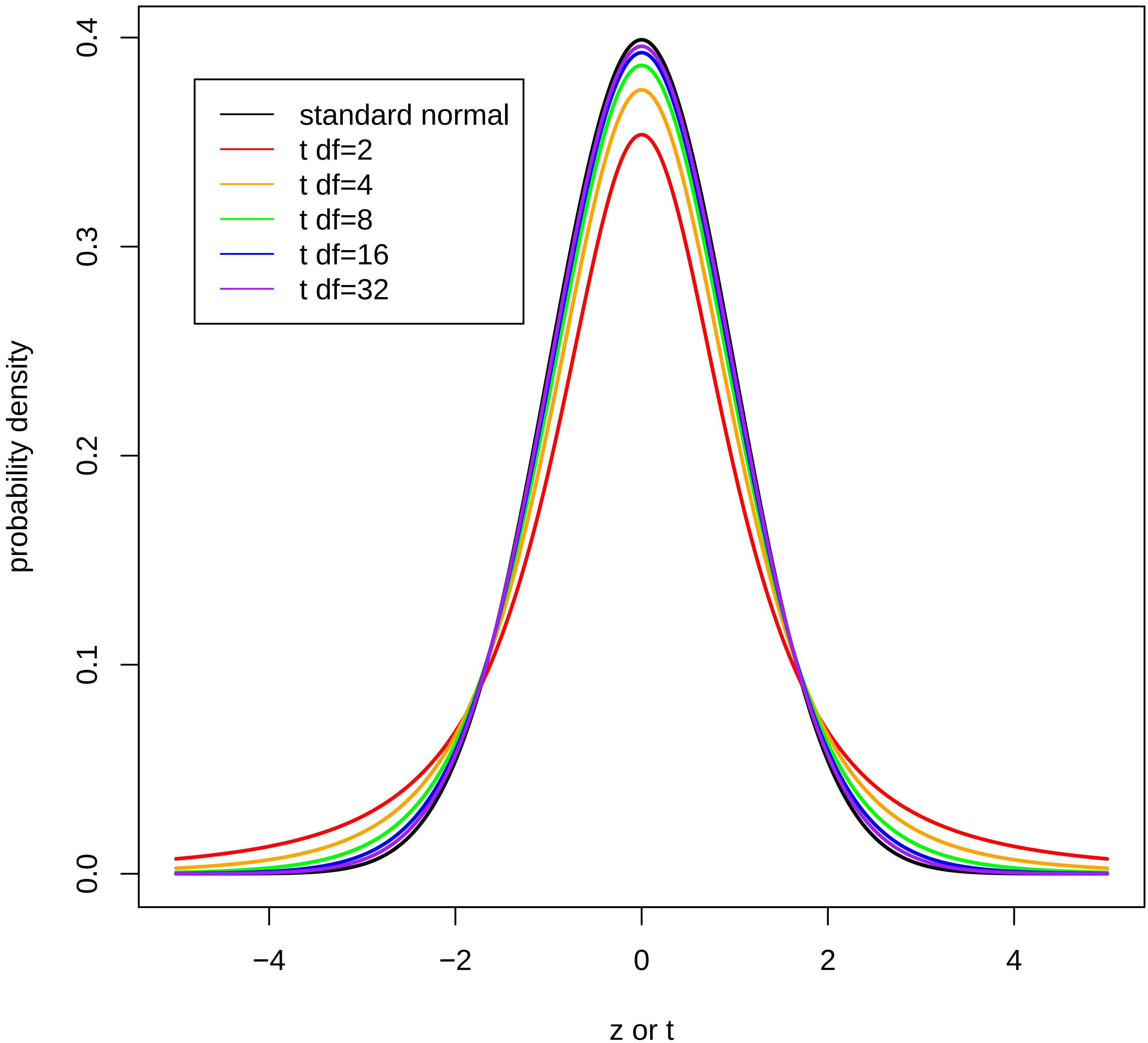


standard normal
Student's t df=4
probability density
z or t
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4



standard normal
Student's t df=4
Pr(|t| >= 1.9600) = 0.1216
Pr(|z| >= −1.9600) = 0.0500
probability density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
z or t


Pr(|t| >= 2.7764) = 0.0500
Pr(|z| >= 2.7764) = 0.0055
standard normal
Student's t df=4
probability density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
z or t


standard normal
t df=2
t df=4
t df=8
t df=16
t df=32
probability density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
z or t

Example (based on Sokal, 1970). Sokal raised *Tribolium castaneum* for 40 generations under a selection regime of killing them adults three days after maturity. Then he measured life span.

In a very large control group, adults lived on average 7.5 days.

We can assume that this is a error-free estimate of the population mean longevity, and the longevity is normally distributed.

Does the population have a different mean intrinsic lifespan after the artificial selection regime?

H_0: the treatment has no effect on longevity.
$\mu = 7.5$days

H_A: $\mu \neq 7.5$days

Step 2: (Random sample) In the experimental lines we observe the following longevity for females

individual #	age at death (days)
1	4.8
2	6.2
3	5.7
4	7.3
5	3.1
6	9.2
7	7.8
8	6.6
9	3.2
10	1.3

Step 3: (Summary stats)
$\bar{a} = 5.52$ $\quad s = 2.42$ $\quad n = 10$

Step 4: (test statistic and P-value)

$$t = \frac{\bar{Y} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{5.52 - 7.5}{\left(\frac{2.42}{\sqrt{10}}\right)} = -2.59$$

Null distribution: Student's t-distribution with 9 degrees of freedom.

$$t_{0.05(2),df=9} = 2.26$$

$$t_{0.02(2),df=9} = 2.82$$

Therefore, $0.02 < P < 0.05$

Step 5: (make a decision) reject the null.

Step 6: We found evidence that evolution in the experimental treatment lowered longevity. The mean longevity for females was 5.52 days (standard deviation = 2.42, $n = 10$), compared to the longevity under the null hypothesis (7.5 days). The difference is too large to be explained by sampling error ($P < 0.05$).

Table 1: Student's t-distribution

df $\boldsymbol{\alpha}$**(2)**	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**	**0.0001**
1	3.08	6.31	12.71	31.82	63.66	636.62	6366.20
2	1.89	2.92	4.30	6.96	9.92	31.60	99.99
3	1.64	2.35	3.18	4.54	5.84	12.92	28.00
4	1.53	2.13	2.78	3.75	4.60	8.61	15.54
5	1.48	2.02	2.57	3.36	4.03	6.87	11.18
6	1.44	1.94	2.45	3.14	3.71	5.96	9.08
7	1.41	1.89	2.36	3.00	3.50	5.41	7.88
8	1.40	1.86	2.31	2.90	3.36	5.04	7.12
9	1.38	1.83	2.26	2.82	3.25	4.78	6.59
10	1.37	1.81	2.23	2.76	3.17	4.59	6.21
11	1.36	1.80	2.20	2.72	3.11	4.44	5.92
12	1.36	1.78	2.18	2.68	3.05	4.32	5.69
13	1.35	1.77	2.16	2.65	3.01	4.22	5.51
14	1.35	1.76	2.14	2.62	2.98	4.14	5.36
15	1.34	1.75	2.13	2.60	2.95	4.07	5.24
16	1.34	1.75	2.12	2.58	2.92	4.01	5.13
17	1.33	1.74	2.11	2.57	2.90	3.97	5.04
18	1.33	1.73	2.10	2.55	2.88	3.92	4.97
19	1.33	1.73	2.09	2.54	2.86	3.88	4.90
20	1.33	1.72	2.09	2.53	2.85	3.85	4.84

Table 2: Student's t-distribution

df $\boldsymbol{\alpha}(\mathbf{2})$	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**	**0.0001**
...							
9	1.38	1.83	2.26	2.82	3.25	4.78	6.59
...							

Estimation:

- point estimate – a best guess of a population parameter based on a sample
- interval estimate – a set of plausible values

Confidence Interval

If we perform many experiments, and construct a 95% confidence interval of the mean for each one, then μ will be contained in $\approx$ 95% of the intervals.

For the 5% of samples that happen to contain the most sampling error:

- we will make Type I errors.
- the 95% confidence interval will exclude μ

95% of the time

$$-t_{0.05(2),df} < \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} < t_{0.05(2),df}$$

therefore 95% of the time:

$$-SE_{\bar{Y}}\left(t_{0.05(2),df}\right) < \bar{Y} - \mu < SE_{\bar{Y}}\left(t_{0.05(2),df}\right)$$

and 95% of the time:

$$SE_{\bar{Y}}\left(t_{0.05(2),df}\right) > -\bar{Y} + \mu > -SE_{\bar{Y}}\left(t_{0.05(2),df}\right)$$

and 95% of the time:

$$\bar{Y} + SE_{\bar{Y}}\left(t_{0.05(2),df}\right) > \mu > \bar{Y} - SE_{\bar{Y}}\left(t_{0.05(2),df}\right)$$

which is the same as:

$$\bar{Y} - SE_{\bar{Y}}\left(t_{0.05(2),df}\right) < \mu < \bar{Y} + SE_{\bar{Y}}\left(t_{0.05(2),df}\right)$$

A $(1-\alpha)100\%$ confidence interval for the mean is

$$\bar{Y} - \left(t_{\alpha(2),df}\right) SE_{\bar{Y}} < \mu < \bar{Y} + \left(t_{\alpha(2),df}\right) SE_{\bar{Y}}$$

The limits of the CI are calculated as:

> the point estimate of the value
>
> $\pm$
>
> (# from a stats table)$\times$(standard error)

For 95% CI, this is like our "2SE rule of thumb" from chapter 4, except we use the appropriate critical value from the t distribution instead of 2.

As n increases, $t_{0.05(2),n-1} \to 1.96$, so the 2SE rule becomes quite accurate.

Beetle example (again):

$\bar{a} = 5.52 \qquad s = 2.42 \qquad n = 10$

$$SE_{\bar{a}} = \frac{s}{\sqrt{n}} = \frac{2.42}{\sqrt{10}} = 0.765$$

$$t_{0.05(2),df=9} = 2.26$$

So we are 95% confident that:

$$5.52 - 2.26 * 0.765 < \mu < 5.52 + 2.26 * 0.765$$

$$3.79 < \mu < 7.24$$

Beetle example (yet again):
$\bar{a} = 5.52 \qquad s = 2.42 \qquad n = 10$

What is the 95% confidence interval for the variance?

$$\left(\frac{df}{\chi^2_{\frac{\alpha}{2},df}}\right) s^2 < \sigma^2 < \left(\frac{df}{\chi^2_{1-\frac{\alpha}{2},df}}\right) s^2$$

$$df = n - 1$$

Beetles, 95% C. I.:

$$\left(\frac{9}{\chi^2_{0.025,9}}\right) 5.86 < \sigma^2 < \left(\frac{9}{\chi^2_{0.975,9}}\right) 5.86$$

$$\left(\frac{9}{19.02}\right) 5.86 < \sigma^2 < \left(\frac{9}{2.70}\right) 5.86$$

$$2.77 < \sigma^2 < 19.5$$

Table 3: The χ^2 distribution

	α									
df	**0.999**	**0.995**	**0.99**	**0.975**	**0.95**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	0.0000016	0.000039	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91
13	2.62	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82	34.53
14	3.04	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32	36.12
15	3.48	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80	37.70
16	3.94	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27	39.25
17	4.42	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72	40.79
18	4.90	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16	42.31
19	5.41	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58	43.82
20	5.92	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00	45.31

Table 4: The χ^2 distribution

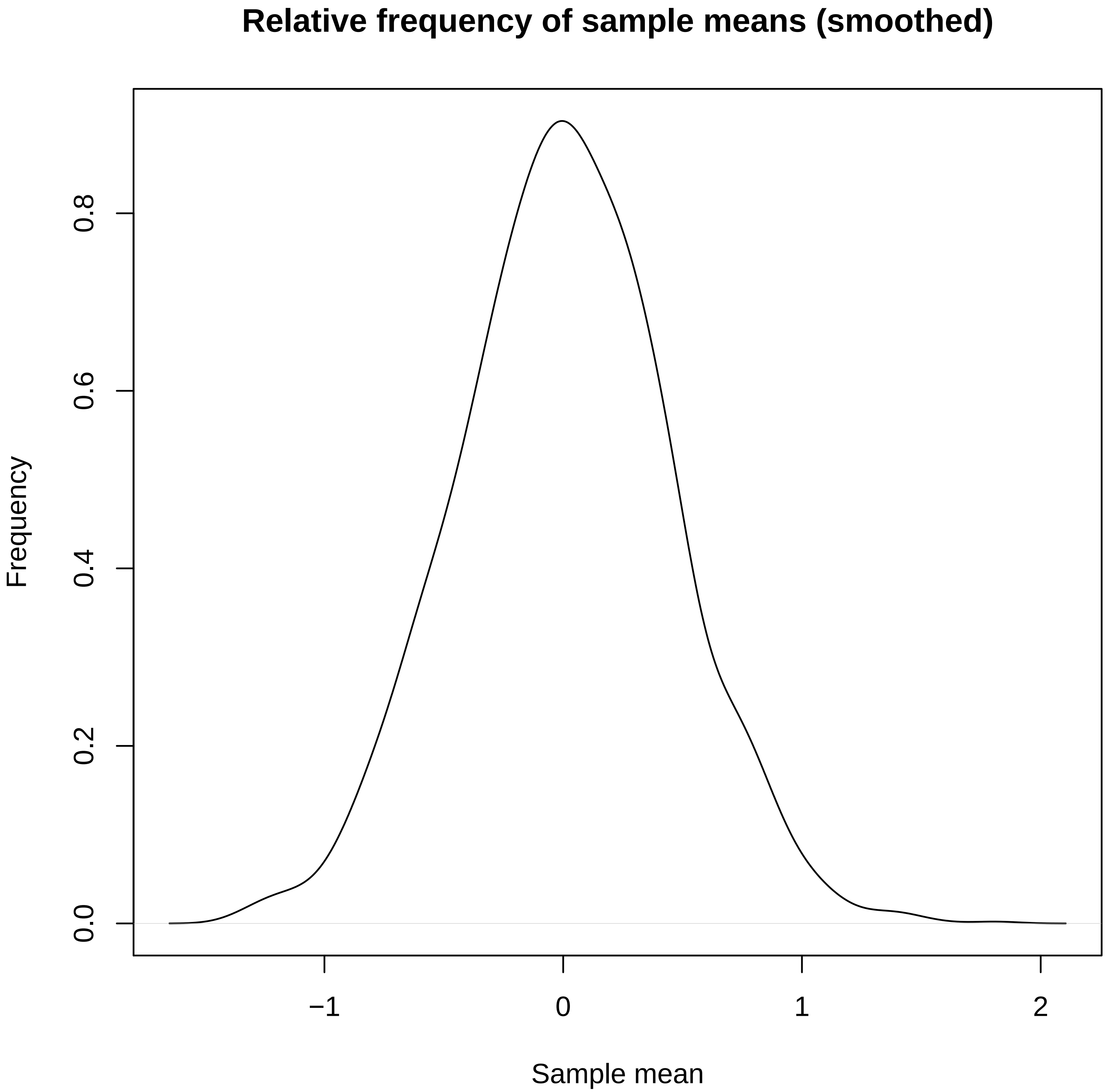
		$\boldsymbol{\alpha}$								
$\boldsymbol{df}$	**0.999**	**0.995**	**0.99**	**0.975**	**0.95**	**0.05**	**0.025**	**0.01**	**0.005**	**0**
. . .										
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	
. . .										

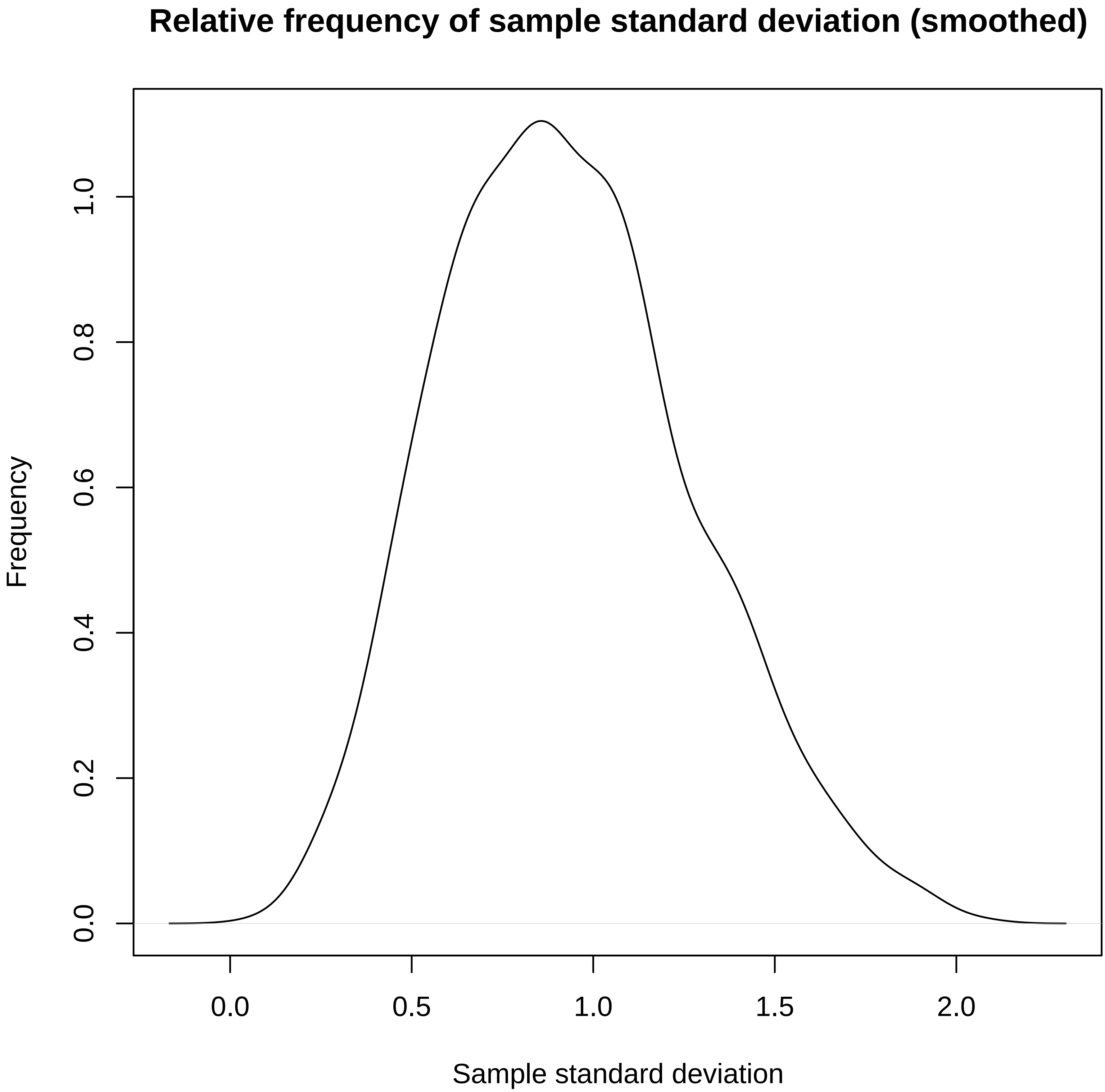
The 95% confidence interval for the standard deviation is found by taking the square root of the variance.

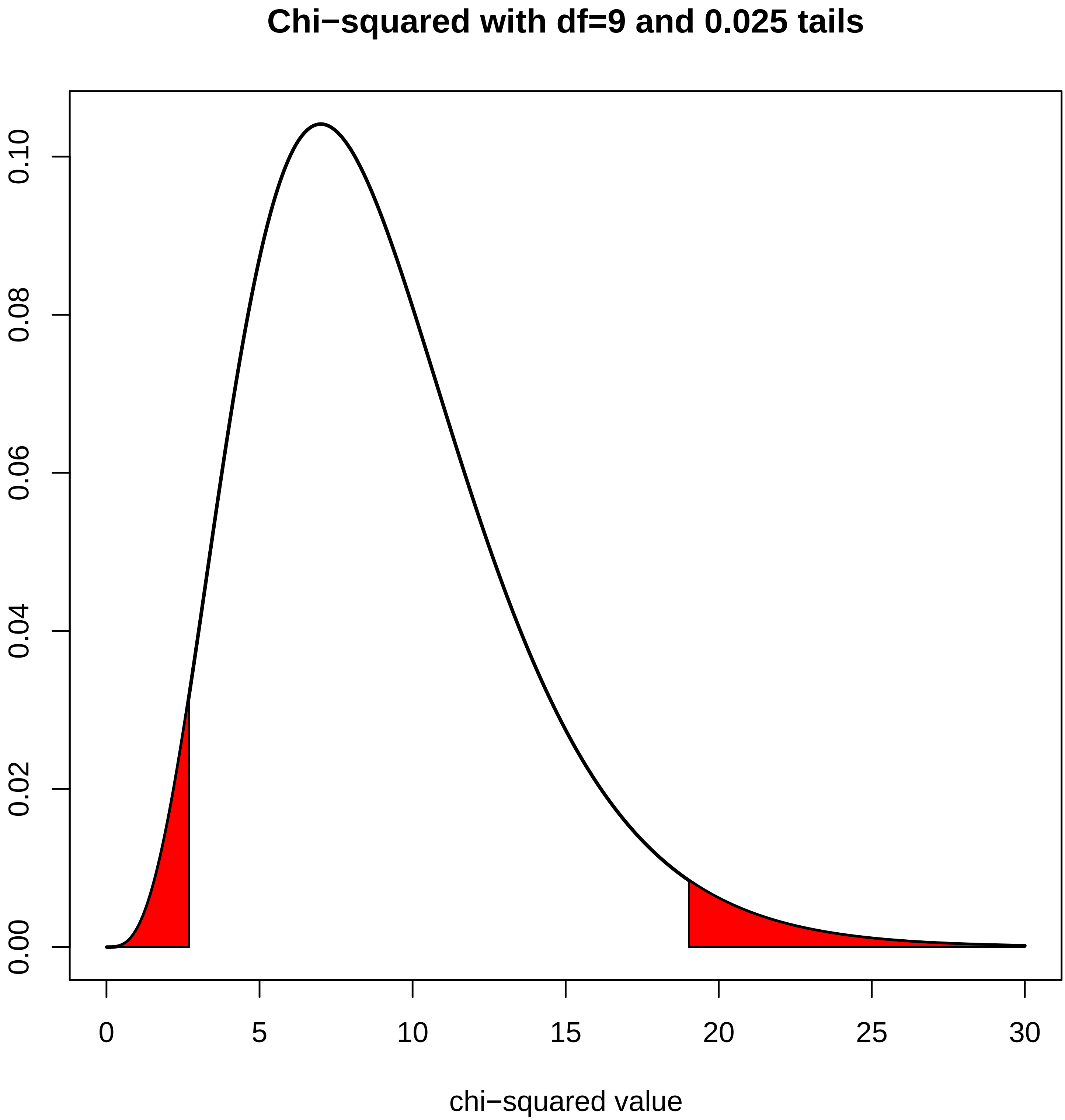
$$2.77 < \sigma^2 < 19.5$$

$$\sqrt{2.77} < \sigma < \sqrt{4.42}$$

$$1.67 < \sigma < 4.42$$


Relative frequency of sample means (smoothed)
Frequency
0.0
0.2
0.4
0.6
0.8
−1
0
1
2
Sample mean


Relative frequency of sample standard deviation (smoothed)
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
Sample standard deviation


Chi−squared with df=9 and 0.025 tails
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
25
30
chi−squared value